

## Enhancement detection distributed denial of service attacks using hybrid n-gram techniques

Andi Maslan<sup>1,2</sup>, Kamaruddin Malik Mohamad<sup>2</sup>, Cik Feresa Mohd Foozy<sup>2</sup>

<sup>1</sup>Department of Informatic Engineering, Universitas Putera Batam, Batam, Indonesia

<sup>2</sup>Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, Batu Pahat Johor, Malaysia

### Article Info

#### Article history:

Received Sep 30, 2020

Revised Apr 28, 2021

Accepted May 15, 2021

#### Keywords:

Chi-square distance

DDoS

Malware

N-grams

Payload

### ABSTRACT

Distributed denial of service (DDoS) attacks have always been a concern of cyber experts. To detect DDoS attacks, several methods can be used. One of the methods used in this research is the n-gram technique. The n-gram approach analyzes the payload of data packets that enter the network to obtain attack patterns. Data is captured and analyzed, after which it is compared with clean data packets. The chi-square distance value close to 1 indicates that the two packages are very similar so that the data packet is not an attack. A deal less than one means the data packet is categorized as an attack. In this research, the threshold for determining the attack level can be lowered to obtain a very high detection accuracy. As a result, the 2-gram technique has a detection accuracy rate with the lowest false positive value of around 13%, with the highest actual positive ratio reaching 99.98%.

*This is an open access article under the [CC BY-SA](#) license.*



### Corresponding Author:

Andi Maslan

Department of Informatics Engineering, Putera Batam University

Muka Kuning Batam, Kepulauan Riau, Batam, Indonesia

Email: Lanmasco@gmail.com

## 1. INTRODUCTION

Distributed denial of service (DDoS) attacks exploit the internet to target critical Web services. This attack aims to prevent unauthorized users from accessing specific network resources or downgrading standard services of legitimate services by sending massive unwanted traffic to the victim (machines or networks) to exhaust connection capacity or bandwidth. This increased flow of DoS attacks has put servers and network devices on the internet at greater risk. The DDoS attack has been for several years. Previous single source attacks are currently countered simply by several defense mechanisms, and therefore the source of those attacks will be rejected or blocked by improved tracing capabilities. However, with the incredible growth of the internet lately, many systems are currently vulnerable to attackers. Attackers will now use a vast range of hosts to launch attacks on the server. Until now, there is not an accurate and fast technique to overcome them [1]. Personal data or company data need to be protected. In securing data, companies usually use cloud systems or create their servers to store data. But when the computer is always connected to the internet, there are also weaknesses. This weakness usually lies in the server's ability to serve all kinds of data access. There are so many requests for data access to the server that it is difficult to control. Thus, an attacker intentionally or unintentionally tries to bring down the server's defense system. With these conditions, it is costly to recover them. Companies usually use intrusion detection systems (IDS) as a line of defense to protect the system from being attacked. The IDS system guarantees that a server protected by IDS certainly has a high level of security because it helps the firewall work. Attacks that fail to be detected by the firewall will filter again on ID [2]. The usual type of attack on the server is the DDoS attack. Based on data obtained, the cybersecurity research and company Kaspersky noted that cyberattacks in DDoS in the first quarter (Q1)

2020 increased sharply compared to the same period in the previous year. The surge in cyber-DDoS attacks occurred on sites related to education and cities. DDoS is a cyber-attack carried out by flooding internet network traffic on servers, systems, or networks. Typically, this attack uses multiple host computers until the target computer cannot be accessed. DDoS itself is a viral attack used by hackers. Kaspersky explained the surge in DDoS attacks in the first three months of 2020 could be because hackers took advantage of situations when people had to do activities at home and were very dependent on digital resources. The coronavirus pandemic [3], which began in the first quarter of 2020, has caused almost all activities, be it studying, working, or relaxing, to shift online. The increasing demand for online resources is well known to cyber attackers, who carry out attacks on the most vital or increasingly popular digital services. The US government's Department of Health and Human Services, several hospitals in Paris, and online game servers are examples of DDoS attacks in February and March. In the DDoS attack report in Q1 2020, Kaspersky also revealed significant growth in attacks on educational resources and the city's official website. In Q1 2020, this number tripled compared to the same period in 2019. The share of such attacks amounted to 19 percent of the total number of incidents in Q1 2020. Kaspersky experts estimate that the increased interest in attackers was because people became more dependent on stable and accessible online resources during the pandemic. If people are increasingly upset about a pandemic and can take preventative action, they might go to an official source of information for the more secure guidance. Many schools and universities have also switched to online learning. In general, the total number of DDoS attacks in Q1 2020 has indeed increased. During this period, Kaspersky DDoS Protection detected and blocked twice the number of attacks compared to Q4 2019, and was 80 percent more than Q1 2019. The average duration of attacks also grew in Q1 2020; DDoS attacks lasted 25 percent longer than Q1 2019. To anticipate the attack, a method or method is needed to detect DDoS attacks. According to researchers, to detect attacks needed a method or algorithm, as well as research conducted by [4]; From a survey, it is found that Naïve Bayes (NB) algorithm provides faster learning/training speed than other machine learning algorithms. Also, it has more accuracy in the classification and detection of attacks. So we are proposing a network intrusion detection system (IDS) that uses a machine learning approach with the help of the NB algorithm. In contrast, the research conducted by [4] the deep learning methodology supported Gaussian-Bernoulli restricted Ludwig Boltzmann machine (RBM) to the detection of denial of service (DoS) attacks is taken into account to extend the DoS attack detection accuracy, seven extra layers area unit additional between the visible and also the hidden layers of the RBM. right, end up in DoS attack detection area unit obtained by optimization of the hyperparameters of the planned deep RBM model. the shape of the RBM that permits application of the continual information is employed. To detect the recent DDoS attack, a researcher [5] said that it needs special treatment at the application layer, such as the hypertext transfer protocol (HTTP), which must be protected. In this protocol, some commands must be considered, such as HTTP requests.

## 2. RESEARCH METHOD

Haris *et al.* [6], projected an excellent defensive system called the web to protect consumer hosts, network routers, and network servers from turning into victims, zombies, and handlers of DDoS flood attacks. It covers any IP-based public network on the internet and uses preventive and rate-limiting to eliminate system vulnerabilities on-track machines. It enforces dynamic security policies for safeguarding network resources against DDoS flood attacks. DDoS instrumentation as a comprehensive framework for DDoS attack detection. It uses a network-based detection technique to defend advanced and leisurely styles of DDoS attacks and works in parallel to examine and manage progress traffic in the period. It covers stateful scrutiny on traffic flow streams and correlates actions among different sessions by continuously observing each DDoS attack and legit applications. It terminates the session when it detects a DDoS attack. Several methods are used to detect DDoS attacks, namely statistical, knowledgebase, soft computing, data mining, and machine learning.

For this research, there are four techniques: statistical-based, knowledge-based, software computing, and machine learning, which in this study is called heuristic detection. This technique was chosen because a payload being analyzed needs to be diagnosed early, such as the number of incoming data packets every second (statistical), incoming packets are standard packets or not (knowledge-based), and to determine whether the level of accuracy of detection of a malicious packet is whether or not it is tested with datasets that are already available on the internet (data mining and machine learning). Details of the techniques used in this study can be seen in Figure 1.

### 2.1. Research framework

To facilitate research, a framework is needed, which explains in detail the process being carried out. The dataset in this study was taken from the Canadian Institute for Cybersecurity (CICIDS2017) and A

management information base (MIB2016) was obtained by capturing packet data using Wireshark, then extracted the data into several features. The packet data are analyzed one by one using the n-gram technique to separate standard packets and packet attacks. N-gram technique can be done using a data payload on a data packet. The research flow in detecting DDoS attacks can be seen in Figure 2.

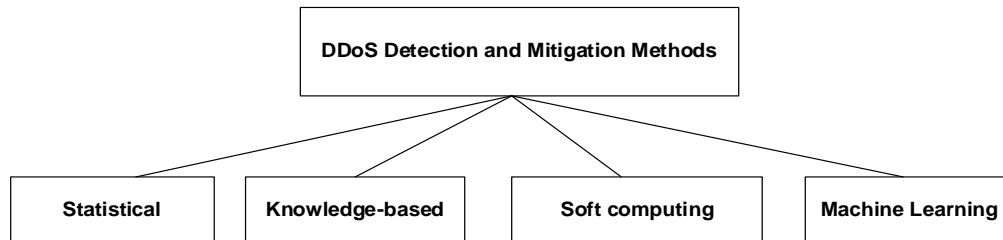


Figure 1. DDoS detection and mitigation methods [7]

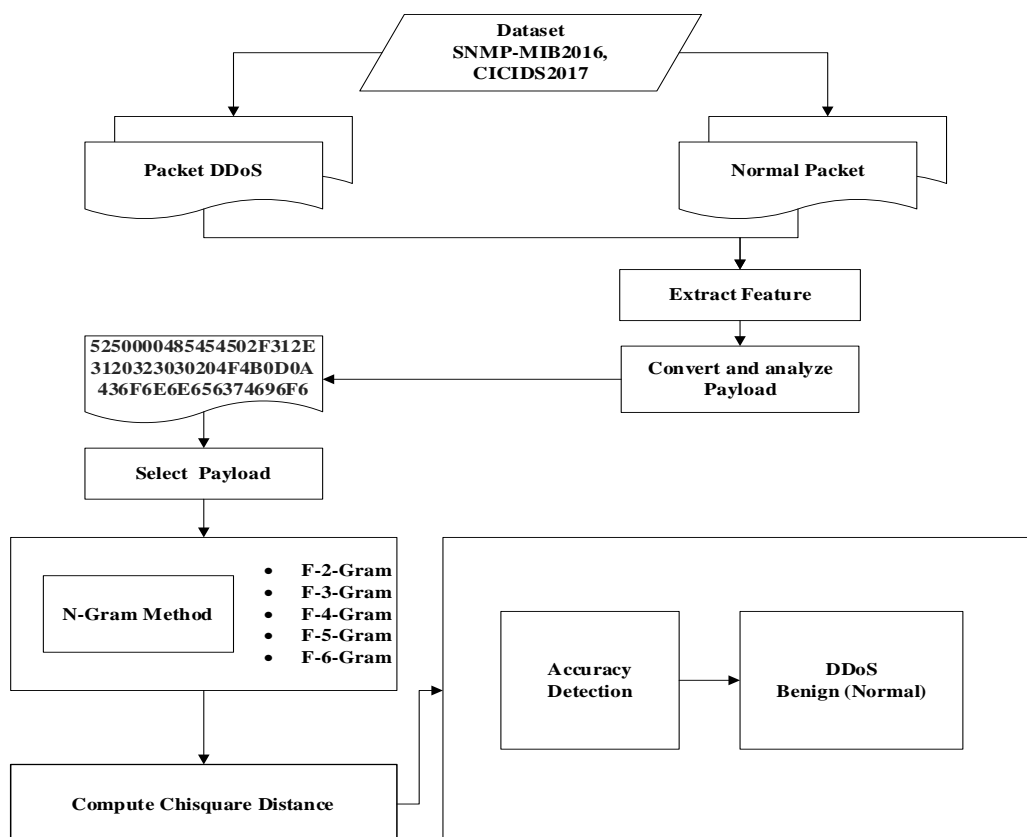


Figure 2. Research framework [8], [9]

The general payload structure is as follows: the data payload is in hexadecimal separated one by one in the form of strings. The formed line will produce a gram sequence starting from 2-gram, 3-gram, 4-gram, 5-gram, and 6-gram. Each gram sequence will determine the number of occurrences of the string (frequency). This applies to two datasets at once, whether it is an attack dataset or a normal (Benign) dataset. It can be called a surveyed package or instruction, or reference package. The number of the regular packet is 529, 918 records. This packet is a comparison to find out the status of the packet being surveyed. To find out the comparative value between packets studied with standard packets, the n-gram technique is used. The number of strings appearing in the n-gram technique aims to calculate the distance in each gram, in this study called chi-square distance. Chi-square distance is a method of finding the distance between two histograms (X and Y). From these calculations, the level of similarity of the packages surveyed with normal packages that have been formed previously can be seen.

## 2.2. Dataset

The selection of the dataset in this study is essential to be able to detect DDoS attacks. To do testing and validation, a dataset is needed so that this research has a maximum contribution. The dataset used in this study was taken from the simple network management protocol-management information base (SNMP-MIB2016) and the Canadian Institute for Cybersecurity (CICIDS2017) protocol and datasets [10]-[12]. This dataset contains two categories of data types, namely normal data and data containing attacks. The data collection period starts on Monday, November 2019, from 9 am to 5 pm in a row for five days. For the first day, the data collected is only in the form of normal data. At the same time, the next day contains data from the brute force file transfer protocol (FTP), brute force secure shell (SSH), DoS, Heartbleed, web attack, infiltration, botnet, and DDoS attacks. Applications used to carry out DoS attacks are Slowloris DoS, DoS slowhttptest, DoS Hulk, and GoldenEye. Meanwhile, a DDoS attack tool call hammer master is used to produce a new dataset called H2NPyalod with a total of 1954 records in the dataset. The accuracy of the new dataset was evaluated to be compared (CICIDS2017) and (SNMP-MIB2016) dataset with 3 algorithms k-nearest neighbor (kNN), neural, and support vector machines (SVM) [10], [13].

## 2.3. Feature extraction

Feature extraction is a technique of taking a feature or feature from a form that later the value obtained will be analyzed for the next process [14]. The features that will be extracted in this study are taken from the results of data packet captured using Wireshark by selecting network protocols such as hypertext transfer protocol (HTTP) and transfer control protocol (TCP), which in this study were taken from 2 datasets available on the internet and one data packet that was captured directly as shown in Figure 3. From Figure 3 [15], the next step is to extract features to get the payload [6] or data in hexadecimal form. Data in hexadecimal form is separated based on data frames with what has been standardized in the TCP/internet protocol (IP) protocol; then, the payload is broken down into strings in this study called n-grams [16], [17].

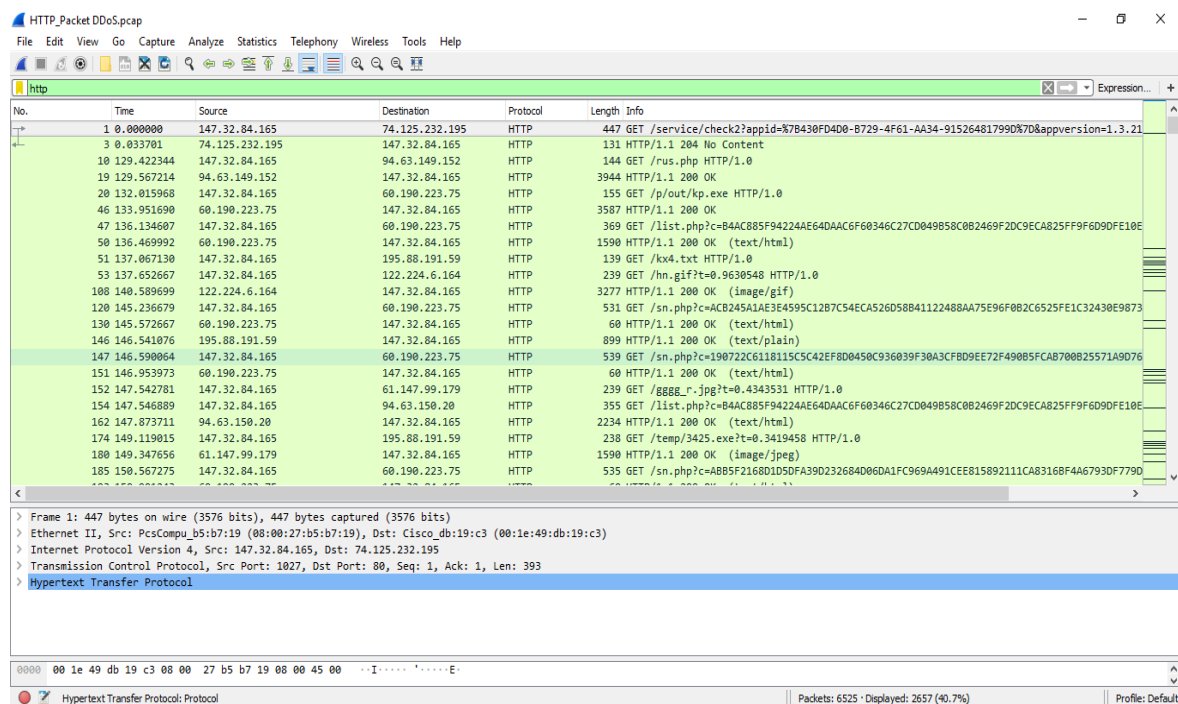


Figure 3. Capture and extract feature

## 2.4. Payload selection

The package payload used in this study is as follows: 192.168.10.15→13.107.4.50 HTTP GET /c/msdownload/update/software/defu/2017/07/am\_delta\_5d55c62f8e8a1e3751fa8dcf66647bb33ae5b343.exe HTTP/1.1.

– Payload Observed

00c1b114eb31001e4fd4ca2808004500016c352a40008006e80dc0a80a0f0d6b0432c12b0050a4f896d828237ace5018010249620000474554202f632f6d73646f776e6c6f61642f7570646174652f736f6674776172652f646565

6752f323031372f30372f616d5f64656c74615f35643535633632663865386131653337353166613864636636  
 3636343762623333616535623334332e65786520485454502f312e310d0a436f6e6e656374696f6e3a204b656  
 5702d416c6976650d0a4163636570743a202a2f2a0d0a4163636570742d456e636f64696e673a206964656e74  
 6974790d0a49662d556e6d6f6469666965642d53696e63653a205765642c203035204a756c203230313720313  
 13a32333a343820474d540d0a52616e67653a2062797465733d31343738382d32343537350d0a557365722d4  
 167656e743a204d6963726f736f667420424954532f372e380d0a486f73743a2061752e646f776e6c6f61642e7  
 7696e646f77737570646174652e636f6d0d0a0d0a

The payload of the structure above is marked in red. The payload is analyzed using n-gram to determine whether a packet is dangerous or not. We then compared with normal packets as marked in the following payload.

– Payload expected

00c1b114eb31b8ac6f1d1f6c08004500008900e440008006c7dac0a80a09451f21e0042100507990c654a5710f  
 3d50180100d4730000474554202f6e6373692e74787420485454502f312e310d0a436f6e6e656374696f6e3a20  
 436c6f73650d0a557365722d4167656e743a204d6963726f736f6674204e4353490d0a486f73743a207777777e  
 6d7366746e6373692e636f6d0d0a0d0a

The sample payload above was analyzed using 3-Gram, with a pattern found between the two payloads that were compared "a20", the analyzed payload was found to be seven patterns. In contrast, the comparison payload was found as many as three patterns. To determine whether a payload is dangerous or not, the chi Square Distance parameter is used, which the results of the analysis in this study are described in full in Table 1 to Table 5.

## 2.5. N-gram technique

N-gram is a sequence of n items from a series of texts or words. This series can be anything, for example, letters, words, or sentences following what we want to use. To find out whether to use an attack or not, the following formula is used:

$$D2(X,Y) = \sum_{i=1}^N \frac{(X_i - Y_i)^2}{Y_i} \quad (1)$$

The value of D is used to test whether the packet analyzed is an attack or not. If the value of chi-square distance is smaller than the value of the chi-square table ( $X_2$ ) or  $\alpha$  value greater than 0.05, then the packet is called an attack/DDoS.

## 2.6. Algorithm classification

To be able to measure the accuracy of DDoS detection with the n-gram technique, the machine learning classification method is used. The algorithm used is supervised learning which consists of kNN, Neural Network, and SVM [18]. The k-Nearest Neighbor algorithm is a supervised algorithm learning where the results of the new instance are classified according to the majority of the k-nearest neighbor categories. The purpose of this algorithm is to classify new objects based on attributes and samples of training data. The k-nearest neighbor algorithm uses the as the predictive value of the new instance value. While the neural networks algorithm adopts the thinking mechanism of a system or application that resembles the human brain, both for processing various received element signals, tolerance for errors. The last algorithm in this study is SVM, a reliable method for solving problems. Data classification. The use of the SVM model processes data into training data and test data. The training data is used in forming the SVM model, while the independent parameter values are selected from the initial data.

---

### Algorithm kNN

---

- 1 For each training pattern <x, f (x)>, add the pattern to the list of training patterns
  - 2 For a pattern, enter Xq
    - For example, x1, x2, ..., xk are k patterns that have the closest distance (neighbors) to xq
    - Return the class that has the most number of patterns among the k patterns as a decision class
- 

---

### Algorithm neural network

---

- 1 Form each pattern pi
    - Wi=pi
    - Form the pattern unit with the input weight vector wi
    - Connect the pattern units to the summing unit for each class
  - End
  - Determine the constant | ck | for each adding unit
-

```

2  For each pattern pi
    K = class pi
    Find the distance, in, to the closest pattern to class k
    Dtot [k] = dtoto [k] + d1
    End
  For each class k
     $\sigma_k = (g. Dtot [k]) / | C_k |$ 
  End

```

#### Algorithm SVM

```

1  Initialization,  $\alpha_i=0$ 
    Calculate the matrix  $D_{ij}=y_i y_j (K(x_i, x_j) + \lambda_2)$ 
2  Perform the three steps below for  $i=1, 1, \dots$ 
    •  $E_i = \sum_{j=1}^l \alpha_j D_{ij}$ 
    •  $\delta \alpha_i = \min\{\max[Y(1-E_i), -\alpha_i], C-\alpha_i\}$ 
    •  $\alpha_i = \alpha_i + \delta \alpha_i$ 
3  Return to Step 2 until  $\alpha$  converges

```

### 3. RESULTS AND ANALYSIS

#### 3.1. Chi-square distance

After extracting the payload data, the payloads are separated using an n-gram algorithm, starting from 2-grams, 3-grams, 4-grams, 5-grams, and 6-grams. Calculating the chi-square distance value for a 2-gram payload shows that the survey payload string has a frequency value of string occurrence. The percentage value is calculated, as well as for standard payload; the percentage is also calculated. After all the steps are done, the chi-square distance value is obtained for each gram; if the chi-square distance value is greater than the 0.9 thresholds, then the payload is abnormal (DDoS), as shown in Table 1.

The survey payload has a string occurrence frequency value while calculating the chi-square distance value for a 3-gram payload. The percentage value is calculated, and for standard payload, the percentage is also calculated. After all the steps are done, the chi-square distance value is obtained for each gram; if the chi-square distance value is greater than the 0.9 thresholds, then the payload is abnormal (DDoS), as shown in Table 2.

The survey payload has a string occurrence frequency value while calculating the chi-square distance value for a 4-gram payload. The percentage value is calculated, and for standard payload, the percentage is also calculated. After all the steps are done, the chi-square distance value is obtained for each gram; if the chi-square distance value is greater than the 0.9 thresholds, then the payload is abnormal (DDoS), as shown in Table 3.

Table 1. 2-Gram analysis payload

Packet Survey				Reference (Normal Packet)		
2-Gram	F	%	Chi-Square Distance	2-Gram	F	Percent
e5	1	0.001322751	0.001203407	71	1	0.00332226
b1	1	0.001322751	0.001203407	b1	1	0.00332226
Chi-Square Distance			0.337496629			

Table 2. 3-Gram analysis payload

Packet Survey				Reference (Normal Packet)		
3-Gram	F	%	Chi-Square distance	3-Gram	F	%
00c	1	0.00132626	0.017435428	0c1	1	0.020000
735	2	0.00265252	0.002417006	0a4	2	0.006667
a20	7	0.00397878	0.00108371	a20	3	0.006667
Chi-Square Distance			0.796257719			

Table 3. 4-Gram analysis payload

Packet Survey				Reference (Normal Packet)		
4-Gram	F	%	Chi-Square Distance	4-Gram	F	%
00c1	1	0.00132626	0.014175074	0c10	0	0.01672241
b114	1	0.00132626	0.001217892	b114	1	0.00334448
0d0a	9	0.01193634	0		9	0.01193634
Chi-Square Distance			0.887882948			

The survey payload has a string occurrence frequency value while calculating the chi-square distance value for a 5-gram payload. The percentage value is calculated, and for standard payload, the percentage is also calculated. After all the steps are done, the chi-square distance value is obtained for each gram; if the chi-square distance value is greater than the 0.9 thresholds [19], then the payload is abnormal (DDoS), as shown in Table 4.

The survey payload has a string occurrence frequency value while calculating the chi-square distance value for a 6-gram payload. The percentage value is calculated, and for standard payload, the percentage is also calculated. After all the steps are done, the chi-square distance value is obtained for each gram; if the chi-square distance value is greater than the 0.9 thresholds, then the payload is abnormal (DDoS), as shown in Table 5.

Table 4. 5-Gram analysis payload

Packet Survey				Reference (Normal Packet)		
5-Gram	F	%	Chi-Square Distance	5-Gram	F	%
00c1b	1	0.00132626	0.001227357	00c1b	1	0.0033557
0c1b1	1	0.00132626	0.001227357	0c1b1	1	0.0033557
76e6c	2	0.00265252	0.002454713		3	0.0039787
Chi-Square Distance			0.888606111			

Table 5. 6-Gram analysis payload

Packet Survey				Reference (Normal Packet)		
6-Gram	F	%	Chi-Square Distance	6-Gram	%	%
00c1b1	1	0.00132626	0.0012369	00c1b1	1	0.003367
0c1b11	1	0.00132626	0.0012369	0c1b11	1	0.003367
46f776	2	0.00265252	0.00247379			
Chi-Square Distance			0.92519837			

### 3.2. Performance comparison

Using the n-gram technique, the number of packets detected as DDoS packages was 269, and standard packets were 1685. From the detection of the n-gram approach, testing the accuracy of DDoS attack detection using machine learning methods is carried out. Algorithms for measuring performance are kNN, Neural Network, and SVM, where the analysis results can be seen in the following Table 6.

From the table, the comparison of the accuracy level of the n-gram technique in detecting DDoS attacks for the support vector machine (SVM) algorithm is the 1-gram detection accuracy rate of 99.00%, 2-gram 99.23%, 3-gram 96.5%, 4-gram 97.14%, 5-gram 96.7%, 6-gram 94.07%, while the level of detection accuracy with the neural network algorithm 1-gram is 99.00%, 2-gram 99.98%, 3-gram 96.3%, 4-gram 95.9%, 5-gram 95.5%, and 6-gram 93.3%, for the kNN algorithm 1-gram 98%, 2-gram 99.98%, 3-gram 96.7 %, 4-gram 98.00%, 5-gram 97.8% and 6-gram 96.3%. When compared with other algorithms, the machine learning techniques or methods in this study were significantly superior.

For the Mahalanobis distance algorithm [20]-[22], the number of grams used only focuses on 2-gram and 4-gram. In contrast, the chi-square distance algorithm [23] uses 1-gram to 5-gram, and the Reputation value algorithm for the same n-gram values from 3-Gram to 6-Gram, while for research [24] only uses 2-gram with a different algorithm such as Cosine Similarity where the accuracy rate is up to 65%, Jaccard index 65% and levenshtein distance 80%. Based on the results of the study in Table 6, it is explained that the most dominant n-gram technique in detecting DDoS attacks is 2-Gram and 3-Gram and has the highest level of accuracy, reaching 99.98%. When compared with research, [25]-[27] it only got 98.7%, meaning that there was a significant difference reaching 1.28%. Visually it can be seen in Figure 4.

Table 6. The performance comparison between our approach and methods in

Accuracy N-Gram Technique					Algorithm	
1-G	2-G	3-G	4-G	5-G	6-G	
99.81%	94.7%	99.00%	99.00%	99.00%	94.07%	Mahalanobis distance
	98.7%					Chi-Square Distance
						Reputation values
	65%					Cosine Similarity
	65%					Jaccard Index
	80%	96.5%	97.14%	96.7%	94.07%	Levenshtein Distance
99.00%	99.23%					SVM
99.00%	99.98%					Neural Network
98.00%	99.98%					kNN

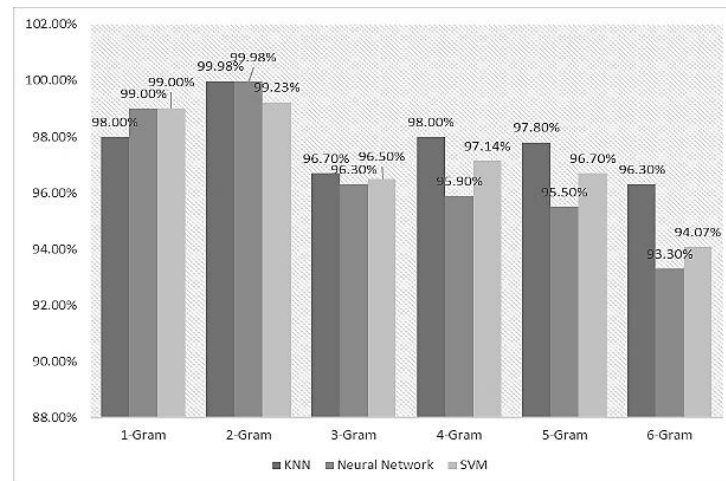


Figure 4. Chart accuracy detection for n-gram technique

#### 4. CONCLUSION

To detect DDoS attacks in this study using n-gram technique. All data packets that flow to the network are captured using tools that have been created before. Applications made have used the Gram technique to separate normal and DDoS packets; all data packets are converted to hexadecimal, the result of this conversion is called the payload. All payloads in the string were analyzed using several n-gram, ranging from 1-gram to 6-gram. The 2-gram and 3-gram techniques have the lowest false positive accuracy rate of 13%, with the highest actual positive ratio reaching a value of 99.98% compared to previous studies that have been done.

#### ACKNOWLEDGEMENTS

We would like to thank Universiti Tun Hussein Onn Malaysia and the University of Putera Batam Indonesian for this research. This work was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme under Grant FRGS/1/2018/ICT04/UTHM/03/4.

#### REFERENCES




- [1] M. Zekri, S. El Kafhali, N. Aboutabit, and Y. Saadi, "DDoS attack detection using machine learning techniques in cloud computing environments," *2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech)*, pp. 1–7, 2018, Feb. 2018, doi: 10.1109/CloudTech.2017.8284731.
- [2] S. Sridharan, "Defeating n-gram Scores for HTTP Attack Detection," *SJSU Scholar Works*, vol. 6, no. San Jose State University, pp. 1–37, 2016, doi: 10.31979/etd.japx-z6eu.
- [3] A. Rizal and A. Hariandy, "Social responsibility of medical journal : a concern for COVID-19 pandemic," *Medical Journal Indonesia*, vol. 29, no. 1, pp. 1–3, 2020, doi: 10.13181/mji.ed.204629.
- [4] Y. Imamverdiyev and F. Abdullayeva, "Deep Learning Method for Denial of Service Attack Detection Based on Restricted Boltzmann Machine," *Big Data*, vol. 6, no. 2, pp. 159–169, Jun. 2018, doi: 10.1089/big.2018.0023.
- [5] P. Prajapati, N. Patel, and P. Shah, "A review of recent detection methods for HTTP ddos attacks," *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 1693–1696, 2019. [Online]. Available: <https://www.ijstr.org/final-print/dec2019/A-Review-Of-Recent-Detection-Methods-For-Http-Ddos-Attacks-.pdf>
- [6] S. H. C. Haris, R. B. Ahmad, and M. A. H. A. Ghani, "Detecting TCP SYN flood attack based on anomaly detection," *2010 Second International Conference on Network Applications, Protocols and Services*, pp. 240–244, Sep. 2010, doi: 10.1109/NETAPPS.2010.50.
- [7] G. S. Kushwah and V. Ranga, "Voting extreme learning machine based distributed denial of service attack detection in cloud computing," *Journal of Information Security and Applications* 53, vol. 53, p. 102532, Apr. 2020, doi: <https://doi.org/10.1016/j.jisa.2020.102532>.
- [8] A. Maslan, K. M. Bin Mohamad, and F. B. Mohd Foozy, "Feature selection for DDoS detection using classification machine learning techniques," *IAES International Journal Artificial Intelligence (IJ-AI)*, vol. 9 No. 1, pp. 137–145, 2020, doi: 10.11591/ijai.v9.i1.pp137-145.
- [9] W. Lee, S. J. Stolfo, and K. W. Mok, "A Data Mining Framework for Building Intrusion Detection Models," *Proceedings of the 1999 IEEE Symposium on Security and Privacy (Cat. No.99CB36344)* pp. 120–132, May 1999, doi: 10.1109/SECPRI.1999.766909.
- [10] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," pp. 108–116, 2018, doi: 10.5220/0006639801080116.
- [11] M. Alkasasbeh, G. Al-Naymat, A. B. A. Hassanat, and M. Almseidin, "Detecting Distributed Denial of Service Attacks Using Data Mining Techniques," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 7, no. 1, 2016, doi: 10.14569/IJACSA.2016.070159.
- [12] F. S. D. L. Filho, F. A. F. Silveira, A. De M. B. Junior, G. V-Solar, and L. F. Silveira, "Smart Detection: An Online Approach for






- DoS/DDoS Attack Detection Using Machine Learning,” *Security and Communication Networks*, vol. 2019, no. Wiley Hindawi, pp. 1–15, 2019, doi: 10.1155/2019/1574749.
- [13] B. Jia, X. Huang, R. Liu, and Y. Ma, “A DDoS Attack Detection Method Based on Hybrid Heterogeneous Multiclassifier Ensemble Learning,” *Journal of Electrical and Computer Engineering*, vol. 2017, no. Hindawi, pp. 1–9, 2017, doi: 10.1155/2017/4975343.
- [14] S. Almutairi, S. Mahfoudh, S. Almutairi, and J. S. Alowibdi, “Hybrid Botnet Detection Based on Host and Network Analysis,” *Journal of Computer Networks and Communications*, vol. 2020, no. Hindawi, pp. 1–16, 2020, doi: 10.1155/2020/9024726.
- [15] A. Maslan, K. M. Mohammad, F. B. M. Foozy, and S. N. Rizki, “DDoS detection on network protocol using neural network with feature extract optimization,” in *Proceedings of ICAITI 2019 - 2nd International Conference on Applied Information Technology and Innovation: Exploring the Future Technology of Applied Information Technology and Innovation*, pp. 92–99, Sep. 2019, doi: 10.1109/ICAITI48442.2019.8982136.
- [16] M. V. Mahoney, “Network traffic anomaly detection based on packet bytes,” *Proc. 2003 ACM Symp. Appl. Comput. - SAC '03*, p. 346, 2003, doi: 10.1145/952589.952601.
- [17] M. Ali, S. Shiaeles, G. Bendiab, and B. Ghita, “Malgra: Machine learning and N-GRAM malware feature extraction and detection system,” *Electronics*, vol. 9, no. 11, pp. 1–20, 2020, doi: 10.3390/electronics9111777.
- [18] M. Aamir and S. M. A. Zaidi, “Clustering based semi-supervised machine learning for DDoS attack classification,” *information sciences*, vol. 33, no. 4, pp. 436–446, May 2021, doi: 10.1016/j.jksuci.2019.02.003.
- [19] B. A. Pramoto and R. M. Ijtihadie, “Sistem Deteksi Intrusi Menggunakan N-Gram Dan Cosine Similarity,” *JUTI Jurnal Ilmiah Teknologi Informasi*, vol. 14, no. 1, p. 108, 2016, doi: 10.12962/j24068535.v14i1.a516.
- [20] F. Angiulli, L. Argento, and A. Furfaro, “Exploiting n-gram location for intrusion detection,” *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, vol. 3, pp. 1–6, 2016, doi: 10.1109/ICTAI.2015.155.
- [21] K. Rieck, and P. Laskov, “Language Models for Detection of Unknown Attacks in Network Traffic Unknown Attacks in Network Traffic,” *Journal in Computer Virology*, vol. 2, no. 4, pp. 243–256, Feb. 2017, doi: 10.1007/s11416-006-0030-0.
- [22] N. Idika, “A Survey of Malware Detection Techniques,” *Purdue Univ.*, p. 48, 2007, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.4594&rep=rep1&type=pdf>.
- [23] A. Oza, K. Ross, R. M. Low, and M. Stamp, “HTTP attack detection using n-gram analysis,” *Computers & Security*, vol. 45, pp. 242–254, Sep. 2014, doi: 10.1016/j.cose.2014.06.002.
- [24] M. Aldwairi, W. Mardini, and A. Alhowaide, “Anomaly payload signature generation system based on efficient tokenization methodology,” *International Journal on Communications Antenna and Propagation (IRECAP)*, vol. 8, no. 5, pp. 421–429, 2018, doi: 10.15866/irecap.v8i5.12794.
- [25] G. Farahani, “Feature Selection Based on Cross-Correlation for the Intrusion Detection System,” *Security and Communication Networks*, Sep. 2020, doi: 10.1155/2020/8875404.
- [26] K. Bouzoubaa, Y. Taher, and B. Nsiri, “Predicting DOS-DDOS Attacks: Review and Evaluation Study of Feature Selection Methods based on Wrapper Process,” *Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 5, pp. 132–145, 2021, doi: 10.14569/IJACSA.2021.0120517.
- [27] A. Manna and M. Alkasasbeh, “Detecting network anomalies using machine learning and SNMP-MIB dataset with IP group,” *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, 2019, doi: 10.1109/ICTCS.2019.8923043.

## BIOGRAPHIES OF AUTHORS






**Andi Maslan**    received the degree in Informatics Engineering was taken at the Budi Utomo Institute of Technology Jakarta (2004) and the master's degree in Computer Science (Information System) completed at the STMIK Putera Batam (2011). Currently, he is finishing his doctoral education at Tun Onn Hussein University Malaysia. The author is a lecturer at the University of Putera Batam and has a functional position as assistant professor. His current research interests include networking, Network Security, and artificial intelligence. He can be contacted at email: Lanmasco@gmail.com.



**Kamaruddin Malik Bin Mohamad**    received the degree in Computer Science and master's degree in computer science (information security) from Universiti Teknologi Malaysia (UTM), in 1992 and 2003, the Ph.D., degree in information technology from Tun Onn Hussein University Malaysia, in 2011. He started his career as a Lecturer at the Department of Information Security and Web Technology, UTHM, in 2004. His research interest includes File Carving, Steganography, Secure Data Wiping, Digital Forensics Triage, Digital Forensic Analysis, Metadata Visualization and Data Redact. He can be contacted at email: malik@uthm.edu.my.



**Cik Feresa Mohd Foozy**    received the degree in information technology and multimedia and the master's degree in computer science (information security) from Universiti Teknologi Malaysia (UTM), in 2006 and 2009, respectively, and the Ph.D. degree in information security from Universiti Teknikal Malaysia Melaka (UTeM), in 2017. She started her career as a Lecturer at the Department of Information Security and Web Technology, UTHM, in November 2011. She is currently an Active Researcher and has been written and presented a number of papers in conferences and journals. She can be contacted at email: feresa@uthm.edu.my.